CHEMBIOCHEM

Supporting Information

The Extension of a DNA Double Helix by an Additional Watson–Crick Base Pair on the Same Backbone

Pawan Kumar,[a] Pawan K. Sharma,[b] Charlotte S. Madsen,[a] Michael Petersen,[a] and Poul Nielsen*[a]

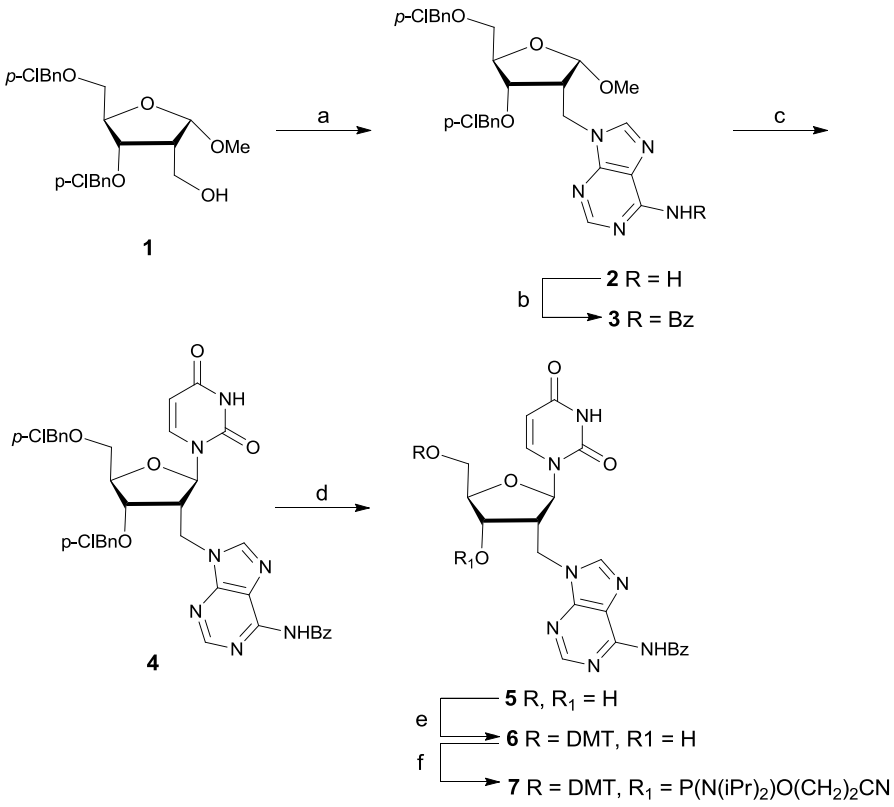
cbic_201300086_sm_miscellaneous_information.pdf

Contents

Synthetic Discussion.

The preparation of the oligonucleotides containing $\mathbf{U_A}$ was accomplished via the the appropriate phosphoramidite. The synthesis of this was performed as shown in Scheme S1. The synthesis started from the known ribose-derivative methyl 3,5-bis-O-(p-chlorobenzyl)-2-deoxy-2-hydroxymethyl-α-D-ribofuranose **1**.[S1] Adenine was introduced as the first nucleobase using a two step procedure involving the activation of the primary hydroxy group of **1** to the triflate, which was subsequently replaced by adenine to generate nucleoside **2**. The presence of $^3J_{CH}$ coupling between the 2'-methylene protons and C-4 as well as C-8 from the adenine in the HMBC NMR spectrum of **2** confirmed the N^9-alkylation. Standard protection of the exocylic amino group afforded a fully protected nucleoside **3**. Vorbrüggen coupling of **3** with silylated uracil was then accomplished in the presence of tin(IV) chloride as the Lewis acid to obtain double-headed nucleoside **4**. Presence of coupling between H-1' and C2 as well as C-6 of the uracil in the HMBC NMR spectrum of **4** confirmed the formation of the double headed nucleoside. The chlorobenzyl groups were removed in a hydrogenation reaction to obtain free double-headed nucleoside **5**. The ROESY spectrum of **5** confirmed the β-configuration as ROE contacts were observed between H6 (uracil) and H2', and between H1' and H4', and between H1' and the methylene protons. Subsequent DMT-protection afforded **6** and phosphitylation provided the phosphoramidite **7**.

[S1] N.-S. Li and J. A. Piccirilli, *J. Org. Chem.* 2004, **69**, 4751.



Scheme S1. Reagents and conditions: (a) i. Tf_2O, pyridine, DCM, ii. Adenine, NaH, DMF, 51%. (b) i. BzCl, pyridine, ii. NH_3 in methanol, 85%. (c) HMDS, $(NH_4)_2SO_4$, CH_3CN, TMSOTf, 58% (d) $Pd(OH)_2/C$, MeOH 41% (e) DMTr Cl, pyridine 48% (f) $NC(CH_2)_2OPClN(iPr)_2$, $EtN(iPr)_2$, CH_2Cl_2 $(ClCH_2)_2$, 65%.

Experimental data for the synthetic preparations

General. All commercial reagents were used as supplied, except CH_2Cl_2 which was distilled prior to use. Anhydrous solvents were dried over 4 Å activated molecular sieves (CH_2Cl_2, DMF, pyridine and, DCE) or 3 Å activated molecular sieves (CH_3CN). All reactions were carried out under an atmosphere of argon or nitrogen, when using anhydrous solvents. Reactions were monitored using TLC analysis with Merck silica gel plates (60 F_{254}). To visualize the plates, they were exposed to UV light (254 nm) and/or immersed in a solution of 5% H_2SO_4 in methanol followed by charring. Standard column chromatography was performed using silica gel 60 (0.040 – 0.063 mm). ^{1}H NMR, ^{13}C NMR and ^{31}P NMR spectra were recorded at 400 MHz, 101 MHz and 162 MHz respectively. Chemical shift values (δ) are reported in ppm relative to either tetramethylsilane (^{1}H NMR) or the deuterated solvents as internal standard for ^{13}C NMR (δ $CDCl_3$ 77.160 ppm, DMSO-d_6 39.52 ppm), and relative to 85% H_3PO_4 as external standard for ^{31}P NMR. 2D spectra (^{1}H-^{1}H COSY, and ^{1}H-^{13}C HSQC) have been used in assigning ^{1}H and ^{13}C NMR signals. High resolution ESI (quadrupole) mass spectra were recorded in positive mode.

Synthesis of 2-*C*-(adenine-9-yl)methyl-3,5-bis-*O*-(4-chlorobenzyl)-2-deoxy-α-D-ribofuranose (2). To a stirred solution of the ribose derivative **1** (2.0 g, 4.68 mmol) in anhydrous CH_2Cl_2 (50 mL) at –70 °C was added dried pyridine (0.5 mL, 6.2 mmol)) and trifluoromethanesulfonic anhydride (1.0 mL, 5.94 mmol). The reaction mixture was allowed to warm to –20°C and stirred for 1 h. The mixture was diluted with CH_2Cl_2 (30 mL) and added H_2O (20 mL) and a saturated aqueous solution of $NaHCO_3$ (10 mL). The phases were separated and the aqeous phase was extracted with CH_2Cl_2 (20 mL). The combined organic phase was dried (Na_2SO_4) and concentrated under reduced pressure. The residue was used without purification in the next step. To a stirred suspension of

adenine (6.3 g, 46.8 mmol) in anhydrous DMF (50 mL) was added NaH (a 60% dispension in oil, 1.31 g, 32.7 mmol). The mixture was stirred at 70 °C for 10 min. and then at rt. A solution of the crude triflate derivative and a catalytic amount of 18-crown-6 (200 mg) in DMF (20 mL) was added. The reaction mixture was stirred at 50°C for 2h and at room temperature for 16 h and then concentrated under reduced pressure. The residue was dissolved in EtOAc (100 mL) and the solution was washed with a saturated aqueous solution of $NaHCO_3$ (50 mL) and brine (50 mL), and the combined aqueous phase was extracted with EtOAc (2 × 50 mL). The combined organic phase was dried (Na_2SO_4) and concentrated under reduced pressure. The residue was purified by column chromatography (0 – 5% MeOH in CH_2Cl_2) to afford the product **2** (1.3 mg, 51%) as a white foam. R_f 0.4 (6% MeOH in CH_2Cl_2); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.34 (s, 1H, H2), 7.73 (s, 1H, H8), 7.31-7.25 (m, 4H, Ar), 7.19 (m, 4H, Ar), 6.14 (s, 2H, NH_2), 4.75 (d, J = 4.9 Hz, 1H, H1'), 4.56 (d, J = 12.3 Hz, 1H, $ArCH_2$), 4.47-4.41 (m, 4H, 2'-CH_2, $ArCH_2$), 4.38 (d, J = 12.3 Hz, 1H, $ArCH_2$), 4.33 (m, 1H, H4'), 3.91 (dd, J = 7.3, 1.8 Hz, 1H, H3'), 3.43 (s, 3H, OCH_3), 3.40 (m, 1H, H5'), 3.32 (dd, J = 9.9, 5.9 Hz, 1H, H5'), 3.03 (m, 1H, H2'); ^{13}C NMR (101 MHz, $CDCl_3$) δ 155.7 (C6), 152.9 (C2), 150.1 (C4), 141.5 (C8), 136.2, 136.2, 133.8, 133.7, 129.3, 129.1, 128.7, 128.7 (Ar), 119.8 (C5), 104.5 (C1'), 83.5 (C4´), 79.1 (C3'), 72.8 ($ArCH_2$), 71.3 ($ArCH_2$), 70.5 (C5'), 55.4 (OCH_3), 46.3 (2'-CH_2), 39.4 (C2'); ESI HRMS: *m/z* 566.1331 ($[M + Na]^+$, $C_{26}H_{27}Cl_2N_5O_4Na^+$ calcd 566.1333).

Synthesis of 2-*C*-(*N*6-benzoyladenine-9-yl)methyl-3,5-bis-*O*-(4-chlorobenzyl)-2-deoxy-α-D-ribofuranose (3). Compound **2** (1.25 mg, 2.29 mmol) was coeporated with anhydrous pyridine (20 mL) and redissolved in the same solvent (20 mL). The solution was stirred at 0°C and added benzoyl chloride (662 μL, 5.72 mmol). The reaction mixture was stirred at room temperature for 16 h and then concentrated under reduced pressure. The residue was coevaporated with toluene (2 × 30 mL) was and then dissolved in MeOH (60 mL) saturated with ammonia. The solution was stirred at

0°C for 30 min. and then concentrated under reduced pressure. The residue was dissolved in EtOAc (60 mL) and the solution was washed with water (2 × 50 mL) and the combined aqueous phase was extracted with EtOAc (2 ×20 mL). The combined organic phase was dried (Na_2SO_4) and concentrated under reduced pressure. The residue was purified by flash column chromatography (0 – 3% MeOH in CH_2Cl_2) to afford the product **3** (1.26 g, 85%) as a white foam. R_f 0.4 (4% MeOH in CH_2Cl_2); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.92 (s, 1H, NH), 8.79 (s, 1H, H2), 8.03 (d, J = 7.3 Hz, 2H, Ar), 7.94 (s, 1H, H8), 7.62 (m, 1H, Ar), 7.55-7.52 (m, 2H, Ar), 7.32-7.30 (m, 4H, Ar), 7.23-7.19 (m, 4H, Ar), 4.77 (d, J = 5.0 Hz, 1H, H1'), 4.57-4.39 (m, 6H, $ArCH_2$, 2'-CH_2), 4.34 (m, 1H, H4'), 3.93 (dd, J = 7.4, 2.2 Hz, 1H, H3'), 3.45 – 3.42 (m, 4H, OCH_3, H5'), 3.35 (dd, J = 10.0, 5.8 Hz, 1H, H5'), 3.05 (m, 1H, H2'); ^{13}C NMR (101 MHz, $CDCl_3$) δ 164.6 (C=O), 152.5, 152.2 (C6, C2), 149.4, 144.0 (Ar, C8), 136.2, 136.0, 133.8, 133.7, 132.8, 132.1, 129.2, 129.0, 128.9, 128.7, 128.6, 128.4, 127.9, 123.2 (Ar, C4, C5), 104.3 (C1'), 83.4 (C4'), 79.0 (C3'), 72.8 ($ArCH_2$), 71.3 ($ArCH_2$), 70.5 (C5'), 55.3 (OCH_3), 46.2 (C2´), 39.5 (2'-CH_2); ESI HRMS: *m/z* 648.1776 ($[M + H]^+$, $C_{33}H_{32}Cl_2N_5O_5^+$ calcd 648.1775).

Synthesis of 2'-*C*-(*N*6-benzoyladenine-9-yl)methyl-3',5'-bis-*O*-(4-chlorobenzyl)-2'-deoxy-uridine (4). A mixture of isonucleoside **3** (0.32 g, 0.5 mmol) and uracil (200 mg, 1.78 mmol) was coevaporated with anhydrous CH_3CN (2 × 10 mL) and then suspended in the same solvent (10 mL). To the stirred suspension was added $(NH_4)_2SO_4$ (8 mg) and HMDS (1.0 mL). The reaction mixture was stirred at 75ºC for 1 h to obtain a clear solution and was then stirred to room temperature and added $SnCl_4$ (25 μL, 0.212 mmol) under argon. The reaction mixture was stirred at room temperature for 16 h and then carefully quenched by the addition of a saturated aqueous solution of $NaHCO_3$ (30 mL). The mixture was filtered through celite and the filter was washed with CH_2Cl_2 (50 mL). The layers were separated and the aqueous phase was extracted with CH_2Cl_2

(20 mL). The combined organic phase was dried ($NaSO_4$) and concentrated under reduced pressure. The residue was purifed by column chromatography (0 – 3% MeOH in CH_2Cl_2) to give the pure β-nucleoside **4** (210 mg, 58%) as a white foam; ^{1}H NMR (400 MHz, $CDCl_3$) δ 11.50 (br s, 1H, NH), 10.07 (s, 1H, NH), 8.71 (s, 1H, H2(A)), 8.09 (s, 1H, H8(A)), 7.91 (d, J = 7.6 Hz, 2H, Ar), 7.45 (m, 1H, Ar), 7.37-7.33 (m, 6H, Ar), 7.28-7.21 (m, 5H, Ar, H6(U)), 6.35 (d, J = 9.0 Hz, 1H, H1´), 5.10 (d, J = 8.0 Hz, 1H, H5(U)), 4.61-4.47 (m, 5H, 2'-CH_2, $ArCH_2$), 4.41 (d, J = 11.8 Hz, 1H, $ArCH_2$), 4.30 (s, 1H, H4'), 4.16 (d, J = 5.3 Hz, 1H, H3'), 3.68 (dd, J = 10.2, 3.0 Hz, 1H, H5'), 3.57-3.39 (m, 2H, H5', H2'); ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.2, 162.8, 152.6, 151.7, 150.6, 150.2, 143.8, 138.7, 135.5, 135.2, 134.3, 134.2, 133.9, 132.4, 129.2, 129.1, 129.0, 128.9, 128.6, 128.2, 122.5 (C=O, Ar, C2,4,5,6,8(A), C2,4,6(U)), 102.8 (C5(U)), 86.9 (C1'), 81.9 (C4'), 80.4 (C3'), 73.1 ($ArCH_2$), 71.0 ((C5'), 70.8 ($ArCH_2$), 45.5 (C2'), 41.2 (2'-CH_2); ESI HRMS: *m/z* 750.1573 ([M + Na]$^+$, $C_{36}H_{31}Cl_2N_7O_6Na^+$ calcd 750.1606).

Synthesis of 2'-*C*-(*N*6-benzoyladenine-9-yl)methyl-2'-deoxyuridine (5). To a stirred solution of compound **4** (440 mg, 0.45 mmol) in MeOH (15 mL) was added $Pd(OH)_2$/C (10-20 wt%, 230 mg), and the suspension was degassed with Argon for 30 min. and then with hydrogen for 1 hour. The mixture was stirred at rt under a hydrogen atmosphere for 35 h. The suspension was filtrated through a layer of celite, which was subsequently washed with warm MeOH (30 mL). The combined filtrates were concentrated under reduced pressure, and the residue was purified by column chromatography (0-12% MeOH) to give the product **5** as a white solid powder (120 mg, 41%) containing ~10% of a compound with a hydrogenated uracil. R_f 0.3 (12% MeOH in CH_2Cl_2). ^{1}H NMR (400 MHz, MeOH-d_4+$CDCl_3$) δ 8.61 (s, 1H, H2(A)), 8.23 (s, 1H, H8(A)), 7.99 (d, J = 8.0 Hz, 2H, Ar), 7.82 (d, J = 8.0 Hz, 1H, H6(U)), 7.55 (m, 1H, Ar), 7.48-7.44 (m, 2H, Ar), 6.18 (d, J = 9.0 Hz, 1H, H1'), 5.53 (d, J = 8.0 Hz, 1H, H5(U)), 4.60 (m, 1H, 2´-CH_2), 4.44 (dd, J = 14.4 Hz, 6.4

Hz, 1H, 2′- CH_2), 4.16 (d, *J* = 5.2 Hz, 1H, H3'), 4.02 (s, 1H, H4'), 3.69 (dd, *J* = 12.0 Hz, 2.8 Hz, H5'), 3.61 (dd, *J* = 12.0 Hz, 2.8 Hz, H5'), 3.00 (m, 1H, H-2'); ^{13}C NMR (101 MHz, MeOH-d_4+$CDCl_3$) δ 167.9, 165.8, 153.3, 152.4, 151.2, 146.0, 142.0, 134.9, 134.0, 129.9, 129.5, 124.4 (Ar, C2,4,5,6,8(A), C2,4,6(U)), 103.6 (C5(U)), 89.1 (C4'), 88.0 (C1'), 73.3 (C3'), 63.3 (C5'), 50.2 (C2'), 41.3 (2′-CH_2); ESI HRMS: *m/z* 480.1631 ($[M - H]^+$, $C_{22}H_{21}N_7O_6.H^+$ calcd. 480.1626).

Synthesis of 5'-O-(4,4'-dimethoxytrityl)-2'-*C*-(*N*6-benzoyladenine-9-yl)methyl-2'-deoxyuridine (6). The double-headed nucleoside **5** (105 mg, 0.21 mmol) was co-evaporated with anhydrous pyridine (2 × 5 mL) and redissolved in the same solvent (5 mL). 4,4'-Dimethoxytrityl chloride (90 mg, 0.26 mmol) was added and the reaction mixture was stirred at room temperature for 16 h. The reaction was quenched by the addition of 2-3 drops of absolute ethanol and the mixture was concentrated under reduced pressure. The residue was dissolved in CH_2Cl_2 (20 mL) and the mixture was washed with a saturated aqueous solution of $NaHCO_3$ (2×10 mL). The combined aqueous phase was extracted with CH_2Cl_2 (2 × 10 mL), and the combined organic phase was dried (Na_2SO_4) and concentrated under reduced pressure. The residue was purified by column chromatography (0-6% MeOH in CH_2Cl_2) to give the product **6** as a pale yellow solid (82 mg, 48%). R_f 0.4 (6% MeOH in CH_2Cl_2). ^{1}H NMR (400 MHz, $CDCl_3$) δ 10.24 (bs, 1H, NH), 9.46 (bs, 1H, NH), 8.78 (s, 1H, H2(A)), 8.27 (s, 1H, H8(A)), 8.01 (d, *J* = 7.6 Hz, 2H, Ar), 7.63 (d, *J* = 8.0 Hz, 1H, H6(U)), 7.57 (t, *J* = 7.6 Hz, 1H, Ar), 7.48 (t, *J* = 7.6 Hz, 2H, Ar), 7.24-7.11 (m, 9H, Ar), 6.76-6.73 (m, 4H, Ar), 6.21 (d, *J* = 9.0 Hz, 1H, H1'), 5.89 (br s, 1H, 3'-OH), 5.37 (d, *J* = 8.0 Hz, 1H, H5(U)), 4.66 (m, 1H, 2'-CH_2), 4.49 (m, 1H, 2'-CH_2), 4.24 (m, 1H, H4'), 4.00 (m, 1H, H3'), 3.73 (s, 3H, OCH_3), 3.72 (s, 3H, OCH_3), 3.48 (dd, *J* = 12.0 Hz, 3.2 Hz, H5'), 3.34 (dd, *J* = 12.0 Hz, 3.2 Hz, H5'), 2.75 (m, 1H, H2'); ^{13}C NMR (101 MHz, DMSO-d_6) δ 165.2, 163.1, 158.9, 152.4, 151.8, 151.2, 150.3, 144.1, 144.2, 139.9, 135.1, 133.5, 133.0, 130.1, 130.2, 128.9, 128.3, 128.2, 128.1, 128.0, 127.3, 123.1, 113.4 (Ar,

C2,4,5,6,8(A), C2,4,6(U)), 103.2 (C5), 87.4 (*C*Ar_3), 86.3 (C4'), 86.2 (C1'), 72.2 (C3'), 64.0 (C5'), 55.4 (OCH_3), 52.0 (C2'), 39.9 (2'-CH_2); ESI HRMS: *m/z* 804.2775 ([M - Na]$^+$, $C_{40}H_{38}N_6O_7Na^+$ calcd. 804.2752).

Synthesis of 5'-*O*-(4,4'-dimethoxytrityl)-3'-*O*-(*P*-β-cyanoethoxy-*N*,*N*-diisopropylaminophosphinyl)-2'-*C*-(*N*6-benzoyladenine-9-yl)methyl-2'-deoxyuridine (7). The nucleoside **6** (78 mg, 0.10 mmol) was co-evaporated with anhydrous 1,2-dichlorethane (2 × 2 mL) and dissolved in anhydrous CH_2Cl_2 (2.0 mL). Diisopropylammonium tetrazolide (40 mg, 0.2 mmol) and 2-cyanoethyl-*N,N,N′,N′*-tetraisopropylphosphoramidite (63 μL, 0.2 mmol) were added, and the mixture was stirred at room temperature for 14 h. Absolute ethanol (2-3 drops) was added, and the reaction mixture was concentrated under reduced pressure. The residue was purified by flash column chromatography (0-4% MeOH in CH_2Cl_2) to give the product **7** (64 mg, 65%) as a white solid. R_f 0.5 (4% MeOH in CH_2Cl_2). ^{31}P NMR (162 MHz, $CDCl_3$) δ 151.1, 149.7. ESI HRMS: *m/z* 1004.3825 ([M - Na]$^+$, $C_{52}H_{56}N_9O_9PNa^+$ calcd. 1004.3831).

Experimental data for oligonucleotide synthesis, hybridisation experiments and CD spectroscopy

Oligonucleotide synthesis. Oligonucleotide synthesis was carried out on an automated DNA synthesizer following the phosphoramidite approach. Synthesis of oligonucleotides **ON1-ON5** (See Table S1) were performed on a 0.2 μmol scale (CPG support) by using the modified nucleoside phosphoramidite **7** to introduce monomer $\mathbf{U_A}$ as well as the corresponding commercial 2-cyanoethyl phosphoramidites of the natural 2′-deoxynucleosides. The synthesis followed the regular protocol for the DNA synthesizer. For the modified phosphoramidite a prolonged coupling time of 15 min. was used. 1*H*-Tetrazole was used as activator. In general coupling yields for the modified phosphoramidite was >50%. For commercial 2-cyanoethyl phosphoramidites of the natural 2'-deoxynucleosides, coupling yield of >95% was obtained. The 5′-*O*-DMT-ON oligonucleotides were removed from the solid support by treatment with concentrated aqueous ammonia at 55°C for 12h. The oligonucleotides were purified by reversed-phase HPLC C18 10μm; 7.8×150 mm column + precolumn: C18 10μm, 7.8×10 mm. Buffer A: 0.05 M Triethyl Ammonium Acetate, pH 7.4. Buffer B: 75 % MeCN/H_2O (3:1, v/v). Program used: 2 min 100% A, 100-30% A over 38 min, 10 min 100% B, 10 min 100% A. All oligonucleotides were detritylated by treatment with 80% aqueous acetic acid for 20 min, neutralized by addition of sodium acetate (3 M, 15 μL), then added sodium perchlorate (5 M, 15 μL) followed by acetone (1 mL). The pure oligonucleotides were precipitated over night at −20°C. The mixture was then placed in a centrifuge and subjected to 12000 rpm, 10 min at 4°C. The supernatant was removed and the pellet washed with cold acetone (2 × 1 mL). The pellet was then dried for 30 min under reduced pressure, dissolved in pure water (1mL) and the concentration measured as OD 260 nm. The purity and constitution of the ON's were confirmed by IC analysis and MALDI-TOF MS $[M - H]^+$, respectively (Table S1).

Thermal denaturation experiments. Samples were dissolved in a medium salt buffer containing 2.5 mM Na_2HPO_4, 5 mM NaH_2PO_4, 100 mM NaCl, and 0.1 mM EDTA at pH = 7.0 with 1.0 μM concentrations of the two complementary oligonucleotide sequences. The increase in UV absorbance at 260 nm as a function of time was recorded while the temperature was increased linearly from 10 to 75°C at a rate of 1.0°C/min by means of a Peltier temperature programmer. The melting curves were found to be reversible. All determinations are averages of at least duplicates within ± 0.5°C.

CD spectroscopy. Samples were dissolved in a medium salt buffer containing 2.5 mM Na_2HPO_4, 5 mM NaH_2PO_4, 100 mM NaCl, and 0.1 mM EDTA at pH = 7.0 with 1.0 μM concentrations of the two complementary oligonucleotide sequences, heated to 80°C and cooled to 10°C. Quartz optical cells with a path length of 5.0 mm were used. CD-spectra (200-350 nm) were recorded on a JASCO J-815 CD-spectrometer as an average of 5 scans using a split of 2.0 nm, and a scan speed of 50 nm/min.

Table S1. MALDI-TOF MS data of synthesized ON's

ON	Sequence	MW (calcd./exp.)
ON1	5′-d(CGC ATA **X**TC GC)-3′	3425.0/3425.0
ON2	5′-d(CGC A**X**A TTC GC)-3′	3425.0/3424.8
ON3	5′-d(GCG AAT A**X**G CG)-3′	3514.0/3515.4
ON4	5′-d(GCG AA**X** ATG CG)-3′	3514.0/3514.7
ON5	5′-d(GCT CAC **X**CT CCC A)-3′	3963.4/3963.0

Table S2. T_m data for modified duplexes.

Entry	Duplex	Zipper	T_m/°C [a] (ΔT_m/°C) [b]
1.	5'-d(CGC AUA ATT CGC) 3'-d(GCG TAT UAA GCG)		46.0
2.	5'-d(CGC A**U$_A$**A TTC GC) 3'-d(GCG TAT UAA GCG)		43.5
3.	5'-d(CGC ATA TTC GC) 3'-d(GCG TAT AAG CG)		45.0
4.	5'-d(CGC ATA **U$_A$**TC GC) 3'-d(GCG TAT AAG CG)		34.0 (–11.0)
5.	5'-d(CGC A**U$_A$**A TTC GC) 3'-d(GCG TAT AAG CG)		40.0 (–5.0)
6.	5'-d(CGC ATA TTC GC) 3'-d(GCG **U$_A$**AT AAG CG)		37.0 (–8.0)
7.	5'-d(CGC ATA TTC GC) 3'-d(GCG TA**U$_A$** AAG CG)		39.5 (–5.5)
8.	5'-d(CGC A**U$_A$**A TTC GC) 3'-d(GCG **U$_A$**AT AAG CG)	(–1)	28.0 (–17.0)
9.	5'-d(CGC ATA **U$_A$**TC GC) 3'-d(GCG TA**U$_A$** AAG CG)	(–1)	28.5 (–16.5)
10.	5'-d(CGC ATA **U$_A$**TC GC) 3'-d(GCG **U$_A$**AT AAG CG)	(–3)	22.0 (–23.0)

[a] Melting temperatures obtained from the maxima of the first derivatives of the melting curves (A_{260} *vs.* temperature) recorded in a buffer containing 2.5 mM Na_2HPO_4, 5.0 mM NaH_2PO_4, 100 mM NaCl, 0.1 mM EDTA, pH 7.0 using 1.0 μM concentrations of each strand. [b]Differences in melting temperatures as compared to the unmodified duplex (entry 3).

Melting curves:

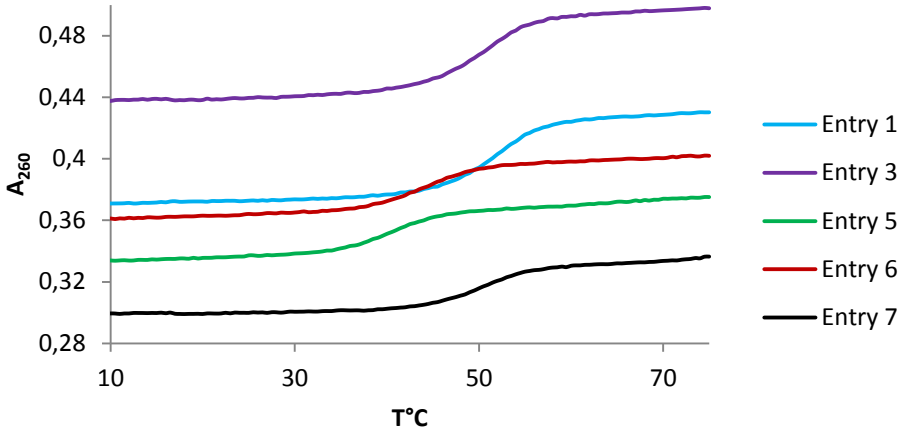


Figure S1. Melting curves for the duplexes given in Table 1.

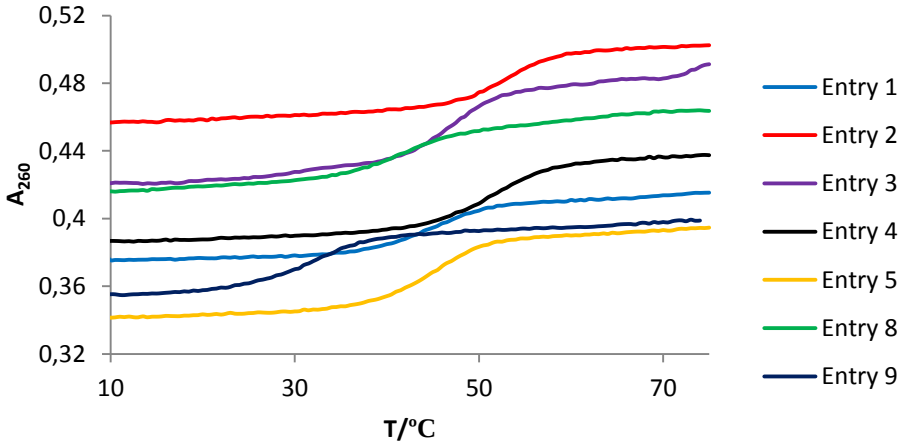


Figure S2. Melting curves for the duplexes given in Table 2.

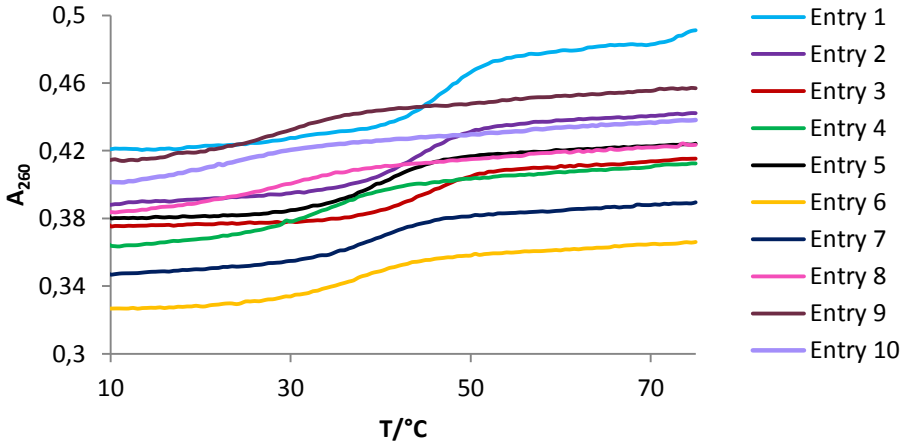


Figure S3. Melting curves for the duplexes given in Table S2.

CD curves:

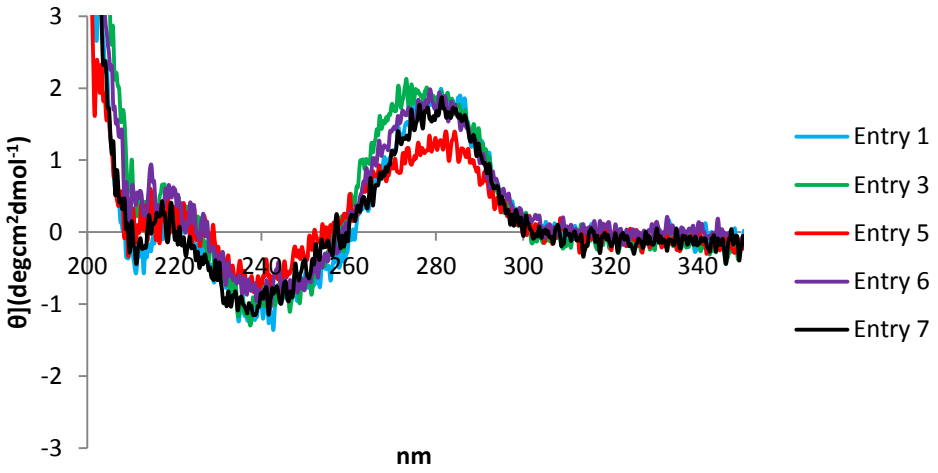


Figure S4. CD curves for the duplexes given in Table 1.

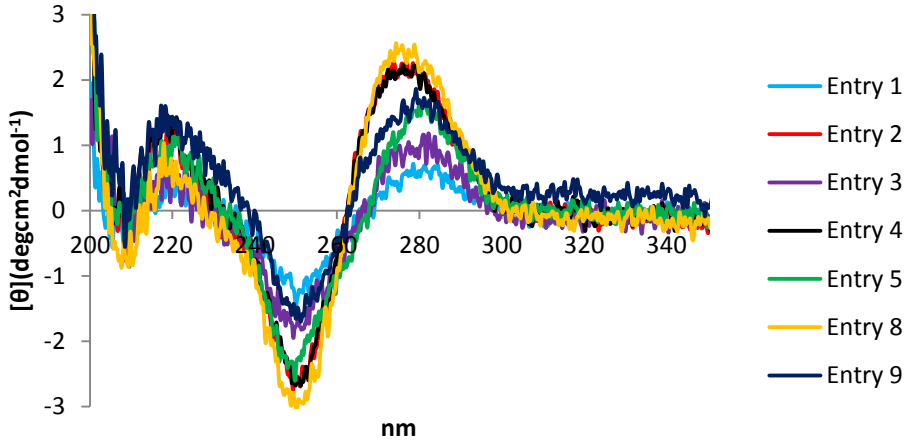


Figure S5. CD curves for the duplexes given in Table 2.

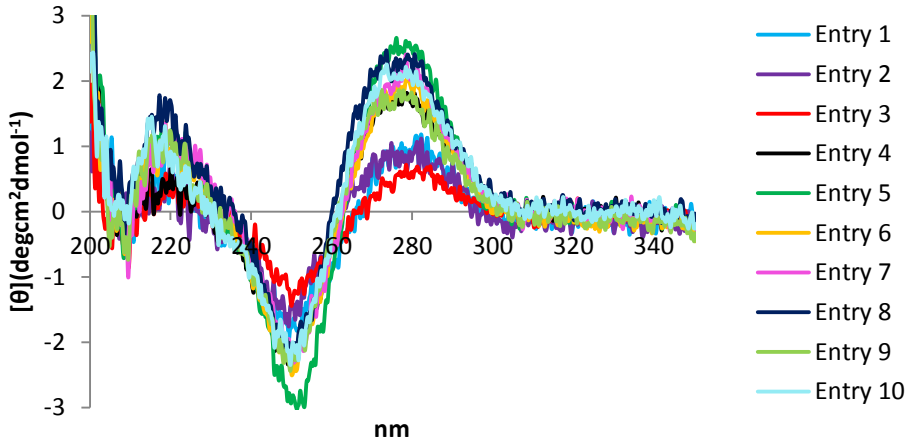


Figure S6. CD curves for the duplexes given in Table S2.

MOLECULAR MODELLING

Setting up the $\boldsymbol{U_A}$ *library file.* The modified nucleoside was built in *xleap* with the two nucleobases stacking and geometry optimised at the HF/6-31G* level using Gaussian03.[S2] The backbone conformation of the optimised nucleoside remained in a duplex-like geometry through the geometry optimisation. Atomic charges were calculated using the RESP methodology[S3] keeping the native charges from the force field of Cornell *et al.* for backbone and sugar atoms except for C1′, C2′, H1′ and H2′.[S4,S5] In this manner we were consistent with the charge derivation for the native atomic charges. Atom types and atomic charges for $\mathbf{U_A}$ are shown in Table S3.

Setting up the deoxy-U library file. The methyl group of thymidine was changed to a hydrogen (atom type HA) in *xleap* and atomic charges were calculated in a manner consistent with the charge derivation in AMBER as described by Kollman and co-workers.[S5]

Building $\boldsymbol{U_A}$*/*$\boldsymbol{U_T}$*-modified duplexes.* All starting coordinates were generated using idealised B-DNA geometries. For duplexes containing $\mathbf{U_A}$ or $\mathbf{U_T}$ residues, helices with two thymidines at the $\mathbf{U_A}$/$\mathbf{U_T}$ site(s) were built initially and the resulting pdb file edited by deleting the second thymidine(s) and the methyl group of the first thymidine(s) and changing the name of the $\mathbf{U_A}$/$\mathbf{U_T}$ unit appropriately to be recognised by the *LEaP* module of AMBER. For duplexes containing deoxy-U residues, helices with a thymidine at the dU were built, the methyl group deleted and the residue name changed appropriately. In *LEaP*, net-neutralizing Na^+ ions were added and the whole system was surrounded by a truncated octahedron of TIP3P waters with a minimum distance of 10.0 Å from the helix to the edges of the box.

MD simulations. All MD simulations were carried out with AMBER9 program suite[S6,S7]using the *ff99*[S8] and *ff99-bsc0*[S9] parameters for nucleic acids and ions. The TIP3P water model was used.[S10] For $\mathbf{U_A}$/$\mathbf{U_T}$-modified systems, an initial energy minimisation was performed with only the $\mathbf{U_A}$/$\mathbf{U_T}$ residue and the 3′-following nucleotide free to move. Harmonic positional restraints with a force constant of 500 kcal mol^{-1} $Å^{-2}$ were applied to the remainder of the system. This energy minimisation relieved the strain on the modified backbone. Afterwards, harmonic positional restraints with a force constant of 500 kcal mol^{-1} $Å^{-2}$ were applied to the entire duplex. The system was minimised for 500 steps of steepest descent and 500 steps of conjugate gradient minimisation.

Finally, a further 1000 steps of steepest descent and 1500 steps of conjugate gradient minimisation was carried out with the entire system free to move.

In the MD equilibration, the SHAKE algorithm was applied with a 1 fs time step. The non-bond cutoff was 9 Å and the particle mesh Ewald method with default parameters was used to calculate long-range electrostatic interactions. The temperature of the system was raised from 0 to 300 K over 20 ps at constant volume using the Berendsen thermostat with a 0.2 ps coupling parameter and applying harmonic positional restraints with a force constant of 10 kcal mol^{-1} $Å^{-2}$ to the DNA molecules. Centre of mass movement was removed every 1000th step. Subsequently, 10 ps MD was carried out at 300 K at constant volume and temperature with parameters as in the first round of equilibration. The final step of equilibration consisted of 200 ps unrestrained *NPT* MD at 300 K with a time step of 2 ps. The reference pressure was 1 atm and the barostat coupling parameter 2.0 ps and the thermostat coupling parameter 1.0 ps. Production runs were carried out for 50 ns with parameters as for the last step of the equilibration except for the pressure and temperature coupling parameters which both were 5.0 ps. For each possible base pairing scheme for the additional A:T base pair, three production runs were initiated with different random initial atom velocities.

Free energy calculations. To evaluate the difference in free energy between the possible base pairing schemes in the 5′-$\mathbf{U_A}$A:3′-A$\mathbf{U_T}$ motif, snapshots where the relevant base pairing was maintained were extracted from the last 40 ns of each trajectories. Snapshots were extracted every 20th ps resulting in 6000 snapshots for Watson-Crick base pairing, 5589 snapshots for reverse Watson-Crick base pairing and 1002 snapshots for reverse Hoogsteen base pairing.

In the MM-PBSA approach, the relevant energy quantities were averaged over the snapshots stripped for explicit water molecules and ions. To calculate the gas phase energy and the solvation free energy, the perl script *mm_pbsa.pl* in the AMBER9 suite was used. The gas phase energy was found from the force field energies using a dielectric constant of 1.0. The polar contribution to the solvation free energy was found by solving the linear Poisson Boltzmann (PB) equation with 1000 steps of iteration. The ionic strength was set to 100 mM, the internal dielectric constant to 1.0 and the external dielectric constant to 80.0. The nonpolar solvation free energy was calculated using the solvent-accessible-surface-area (SASA), a surface tension of 0.0072 kcal $Å^{-2}$ mol^{-1} and no correcting offset. The entropic contribution was estimated using the *nmode* program in AMBER9 assuming the rigid rotor and harmonic approximation to the normal modes at 300 K. For the

calculation of the entropy, the stripped snapshots were prepared in the same way as described above. These structures were minimized for 50,000 steps or until the root-mean-square of the elements of the gradient vector reached below 10^{-4} kcal mol^{-1} Å^{-1} using a distance-dependent dielectric constant of 4r and a non-bonded cut-off of 99 Å and then minimised further with the Newton-Raphson method up to 100 steps until the root-mean-square of the elements of the gradient vector reached below 10^{-5} kcal mol^{-1} Å^{-1}.

S2 M. J. Frisch *et al.* Gaussian 03, Revision D.02 **2004**, Gaussian, Inc., Wallingford, CT.

S3 C. I. Bayly, P. Cieplak, W. D. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269-10280.

S4 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

S5 P. Cieplak, W. D. Cornell, C. Bayly, P. A. Kollman, *J. Comp. Chem.* **1995**, *16*, 1357-1377.

S6 D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham III, S. DeBolt, D. M. Ferguson, G. L. Seibel, P. A. Kollman, *Comp. Phys. Commun.* **1995**, *91*, 1-41.

S7 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J. Comp. Chem.* **2005**, *26*, 1668-1688.

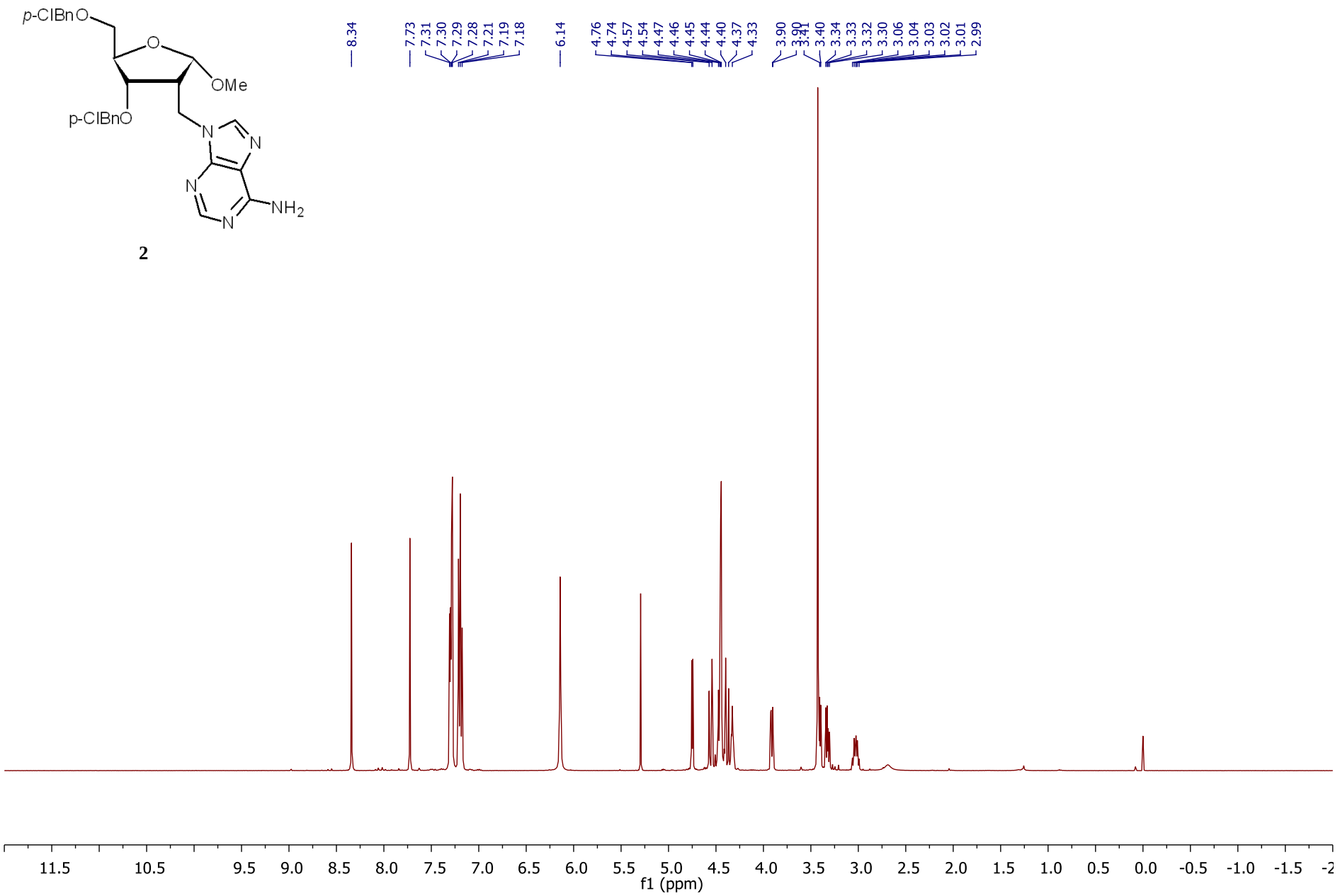
S8 J. M. Wang, P. Cieplak, P. A. Kollman, *J. Comp. Chem.* **2000**, *21*, 1049-1074.

S9 A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, M. Orozco, *Biophys. J.* **2007**, *92*, 3817-3829.

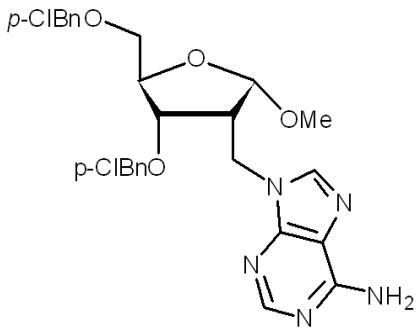
S10 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926-935.

Table S3. Atom types and partial atomic charges for the $\mathbf{U_A}$ nucleotide. C6′, H6′$_a$ and H6′$_b$ are the atoms in the methylene linker and atoms in the additional adenine base are labelled with an "(A)".

Atom Name	Atom type	Partial charge	Atom name	Atom type	Partial charge	Atom name	Atom type	Partial charge
P	P	1.165900	H3′	H1	0.098500	O2	O	−0.568385
O1P	O2	−0.776100	O3′	OS	−0.523200	N9(A)	N*	−0.074524
O2P	O2	−0.776100	C6′	CT	−0.054842	C8(A)	CK	0.093916
O5′	OS	−0.495400	H6′$_a$	H1	0.106284	H8(A)	H5	0.160471
C5′	CI	−0.006900	H6′$_b$	H1	0.106284	N7(A)	NB	−0.536798
H5′$_a$	H1	0.075400	N1	N*	0.131023	C5(A)	CB	0.014810
H5′$_b$	H1	0.075400	C6	CM	−0.114132	C6(A)	CA	0.675204
C4′	CT	0.162900	H6	H4	0.258196	C4(A)	CB	0.423008
H4′	H1	0.117600	C5	CM	−0.423694	N3(A)	NC	−0.659643
O4′	OS	−0.369100	H5	HA	0.192830	N1(A)	NC	−0.730828
C1′	CT	−0.159785	C4	C	0.691706	C2(A)	CQ	0.516247
H1′	H2	0.239636	O4	O	−0.575474	N6(A)	N2	−0.895838
C2′	CT	0.007816	N3	NA	−0.465520	H61(A)	H	0.404126
H2′	HC	0.060566	H3	H	0.328269	H62(A)	H	0.404126
C3′	CT	0.071300	C2	C	0.561750	H2(A)	H5	0.062995

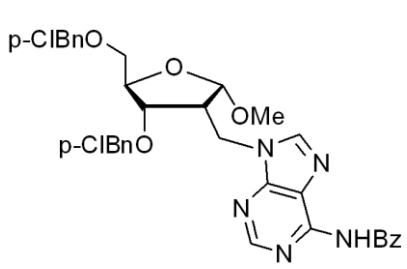

p-ClBnO
O
OMe
p-ClBnO
N
N
N
N
NH2
8.34
7.73
7.31
7.30
7.29
7.28
7.21
7.19
7.18
6.14
4.76
4.74
4.57
4.54
4.47
4.46
4.45
4.44
4.40
4.37
4.33
3.90
3.40
3.34
3.33
3.32
3.30
3.06
3.04
3.03
3.02
3.01
2.99
11.5
10.5
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
f1 (ppm)

2


p-ClBnO
O
OMe
p-ClBnO
N
N
N
N
NH2

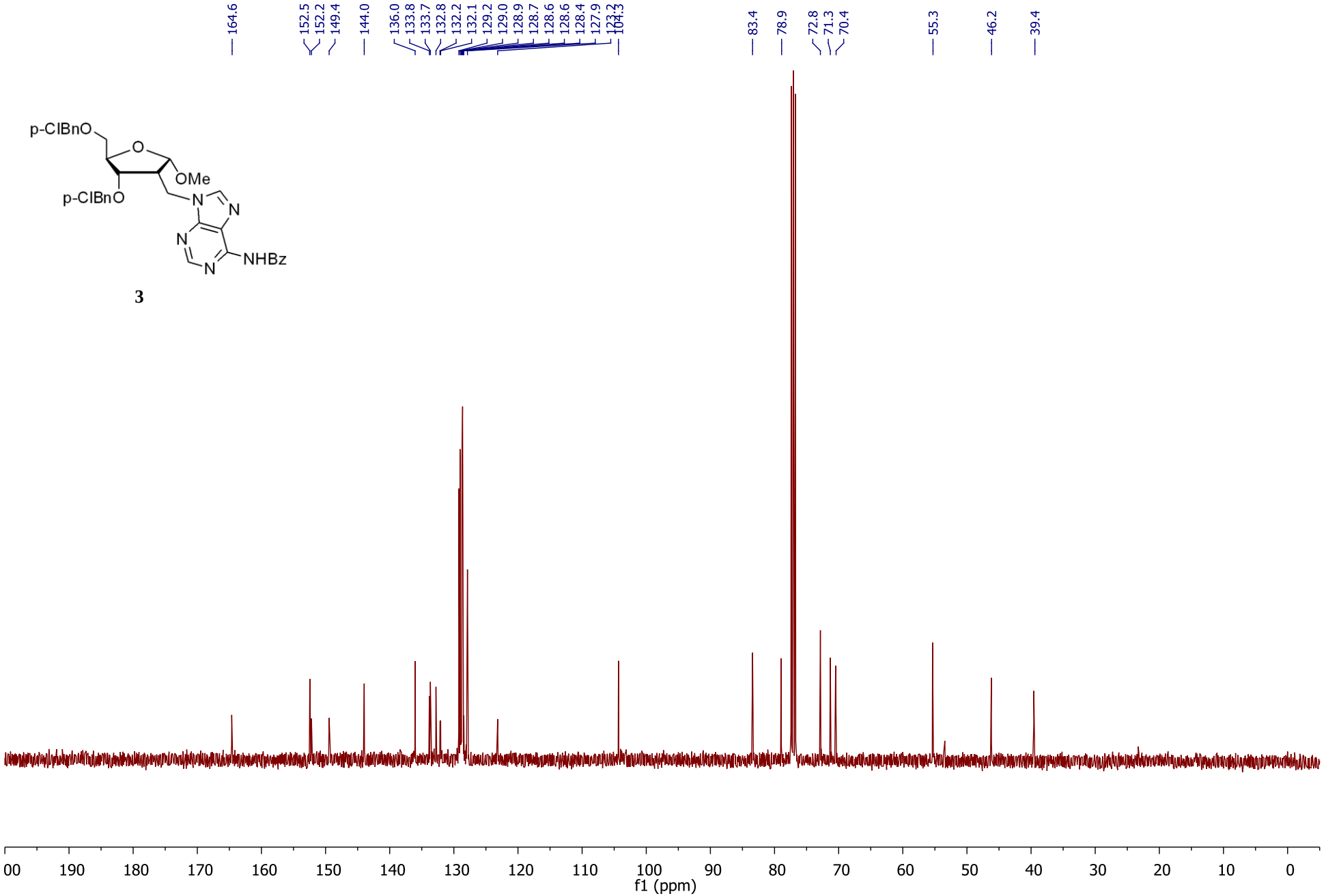
2

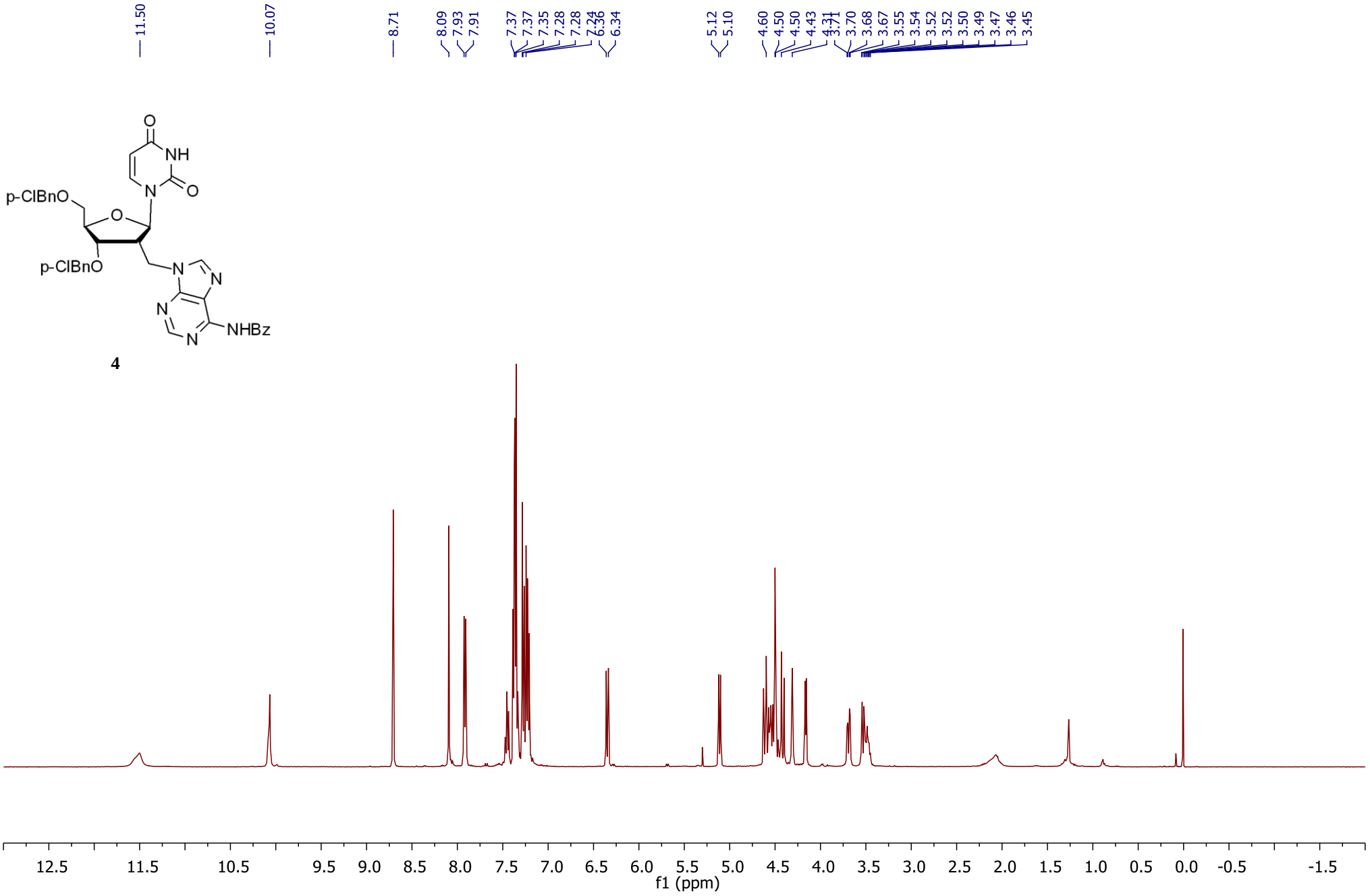
155.7
153.0
150.1
141.5
136.3
136.2
133.8
133.7
129.3
129.1
128.8
128.7
119.8
111.7
104.5
83.6
79.1
77.5
77.2
76.8
72.9
71.4
70.6
55.5
46.3
39.4
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)


p-ClBnO
O
OMe
p-ClBnO
N
N
N
N
NHBz

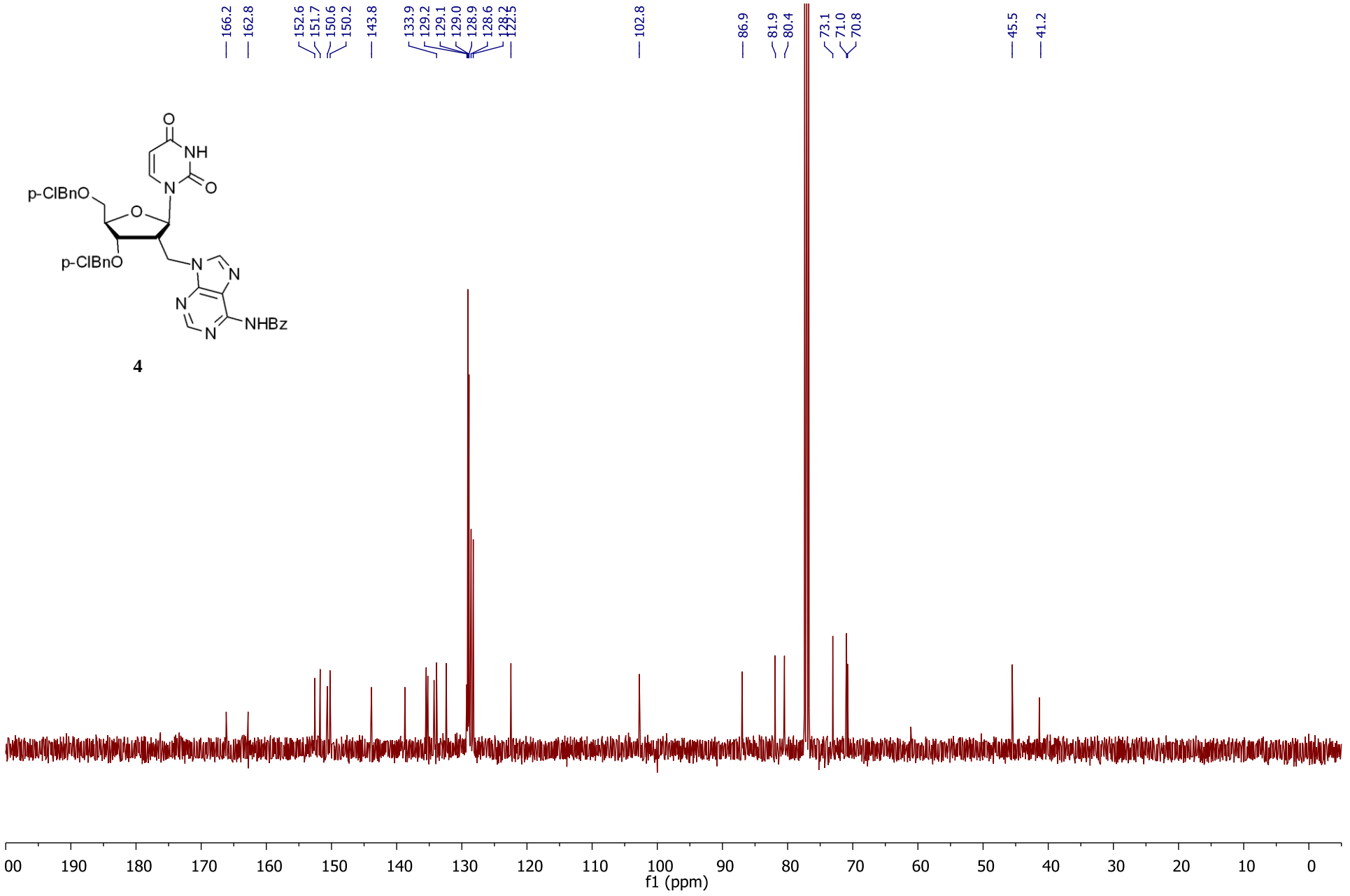
3

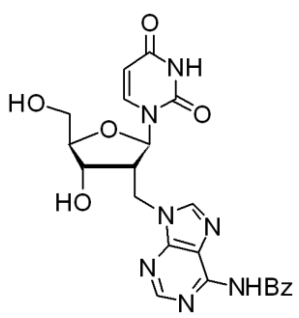
8.92
8.79
8.04
8.02
8.02
7.95
7.64
7.62
7.60
7.55
7.54
7.52
7.32
7.32
7.30
7.30
7.23
7.21
7.19
4.78
4.77
4.59
4.56
4.55
4.54
4.53
4.52
4.46
4.46
4.41
3.92
3.92
3.44
3.43
3.42
3.37
3.35
3.34
3.33
3.09
3.07
3.06
3.05
3.04
3.02
2.0
11.5
11.0
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
f1 (ppm)


164.6
152.5
152.2
149.4
144.0
136.0
133.8
133.7
132.8
132.2
132.1
129.2
129.0
128.9
128.7
128.6
128.6
128.4
127.9
123.2
104.3
83.4
78.9
72.8
71.3
70.4
55.3
46.2
39.4
p-ClBnO
O
OMe
p-ClBnO
N
N
N
N
NHBz
3
00
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)


11.50
10.07
8.71
8.09
7.93
7.91
7.37
7.37
7.35
7.28
7.28
7.24
7.24
6.36
6.34
5.12
5.10
4.60
4.50
4.50
4.43
4.31
4.31
3.70
3.68
3.67
3.55
3.54
3.52
3.52
3.50
3.49
3.47
3.46
3.45
O
NH
N
O
p-ClBnO
O
p-ClBnO
N
N
N
N
NHBz
12.5
11.5
10.5
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.5
f1 (ppm)

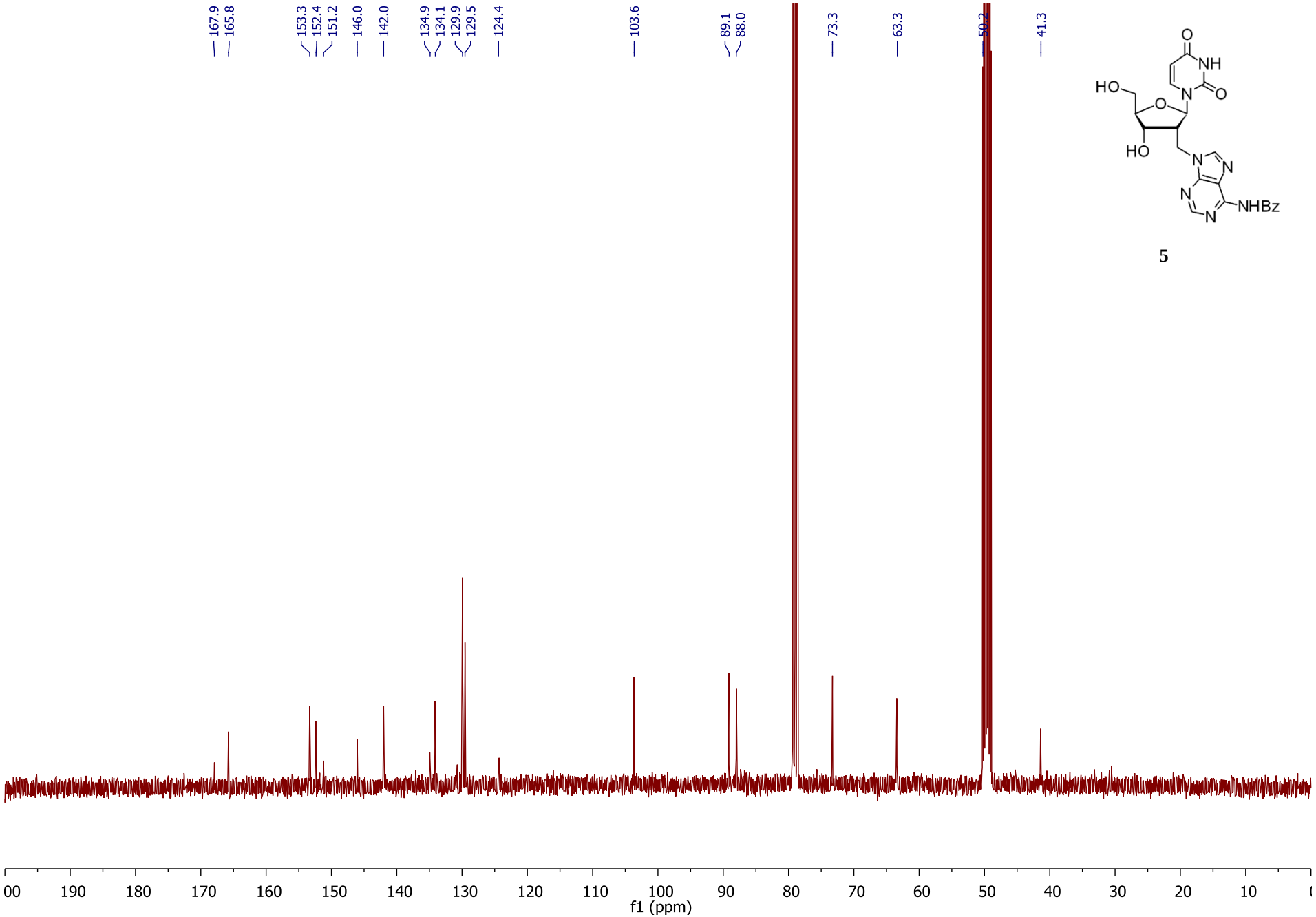
4

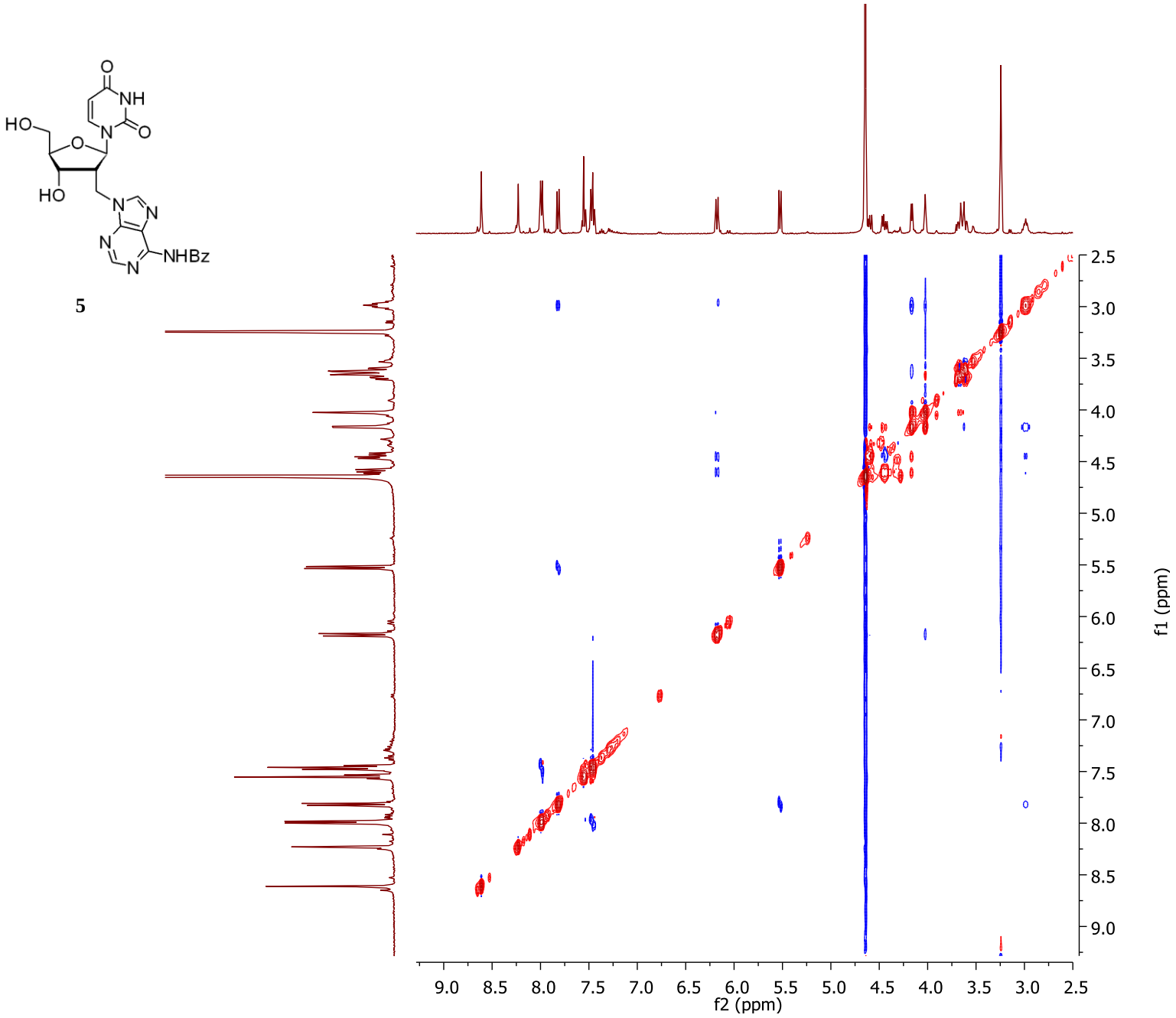

166.2
162.8
152.6
151.7
150.6
150.2
143.8
133.9
129.2
129.1
129.0
128.9
128.6
128.2
122.5
102.8
86.9
81.9
80.4
73.1
71.0
70.8
45.5
41.2
O
NH
N
O
p-ClBnO
O
p-ClBnO
N
N
N
N
NHBz
4
f1 (ppm)


O
NH
HO
N
O
O
HO
N
N
N
N
NHBz

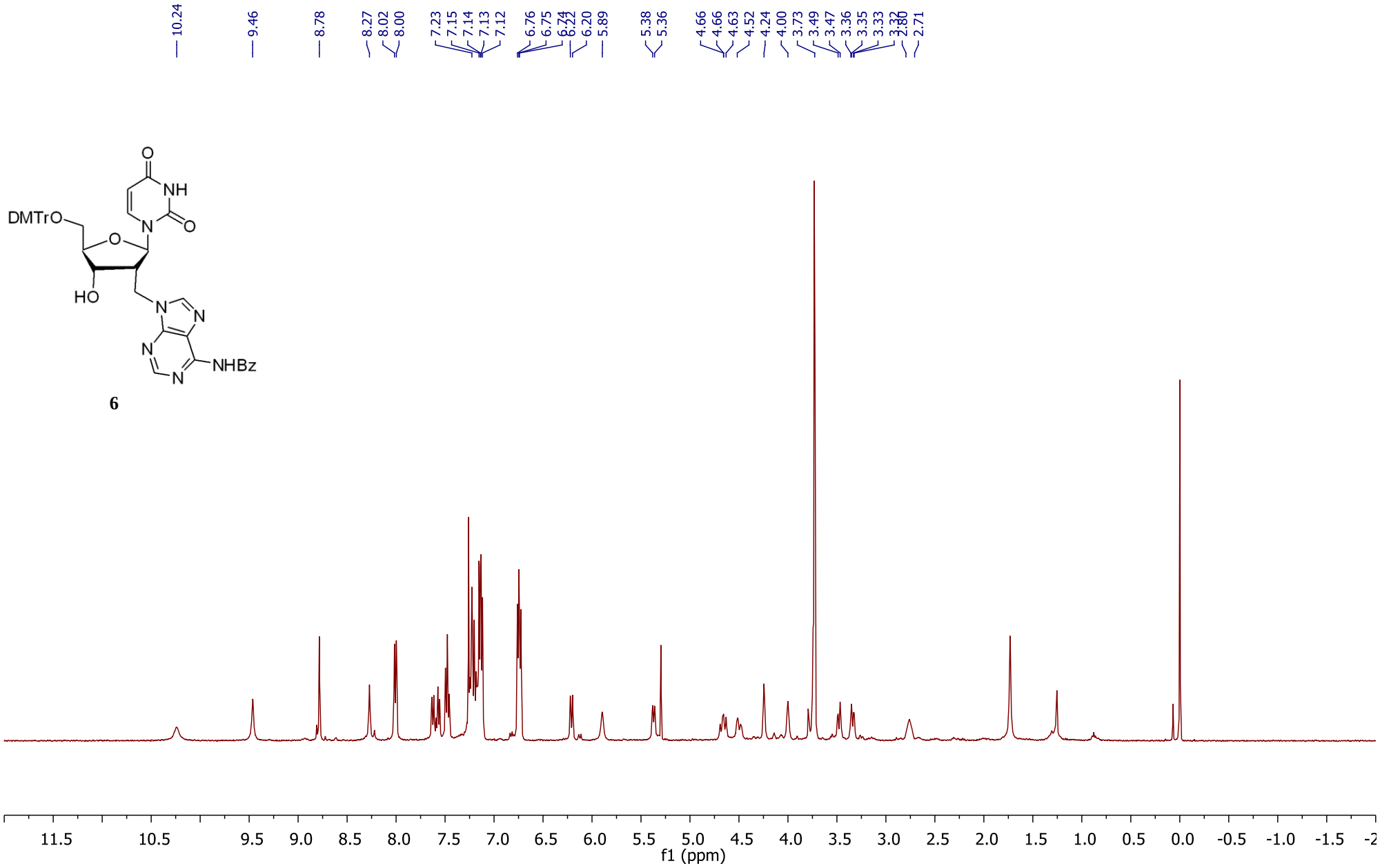
5

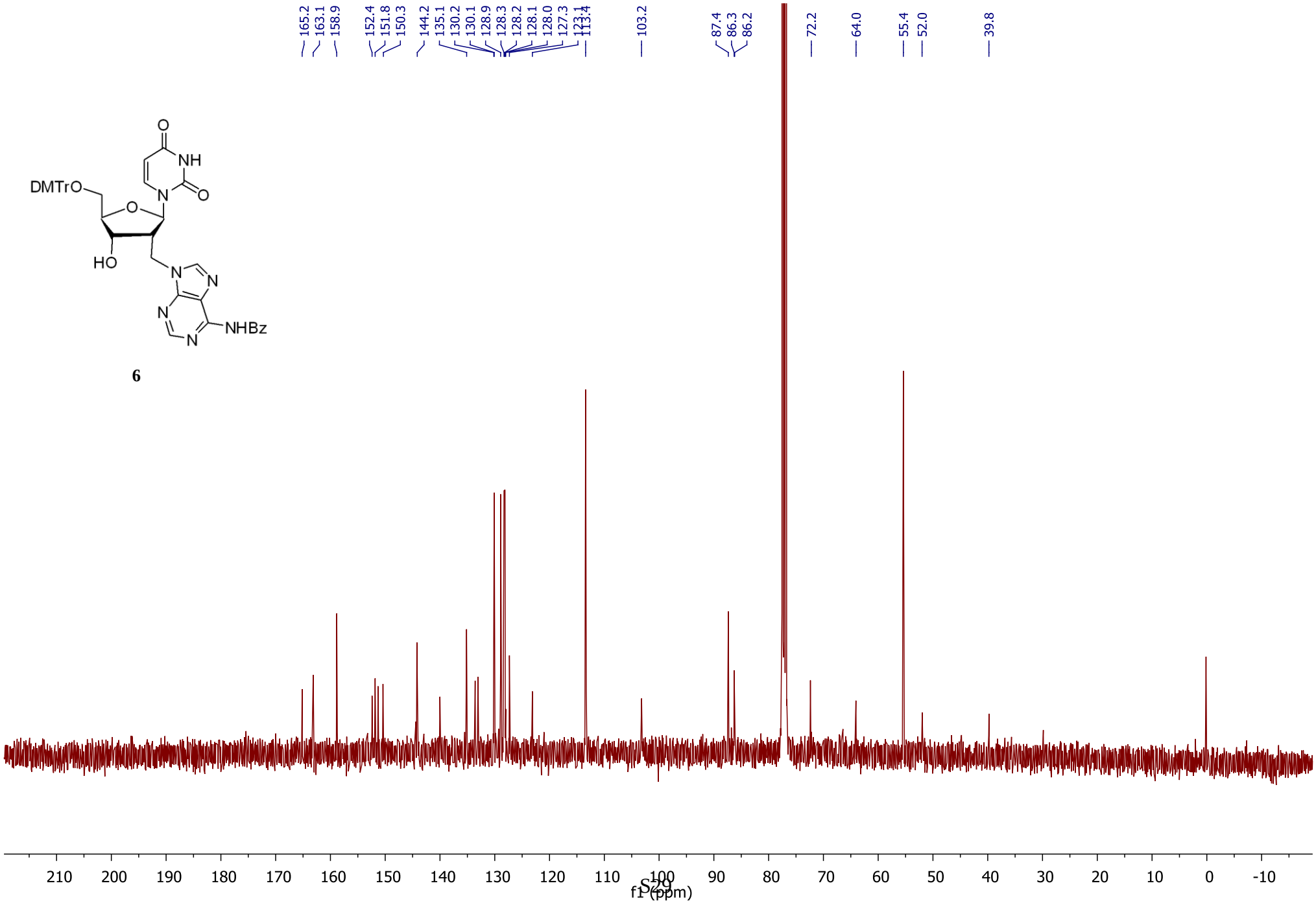
8.61
8.23
8.00
7.98
7.83
7.81
7.57
7.55
7.53
7.48
7.46
7.44
6.19
6.17
5.54
5.52
4.60
4.58
4.47
4.45
4.43
4.42
4.17
4.16
4.02
3.66
3.65
3.63
3.62
3.61
2.95
11.5
10.5
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2
f1 (ppm)


167.9
165.8
153.3
152.4
151.2
146.0
142.0
134.9
134.1
129.9
129.5
124.4
103.6
89.1
88.0
73.3
63.3
50.2
41.3
HO
O
NH
N
O
HO
O
N
N
N
N
NHBz
5
f1 (ppm)

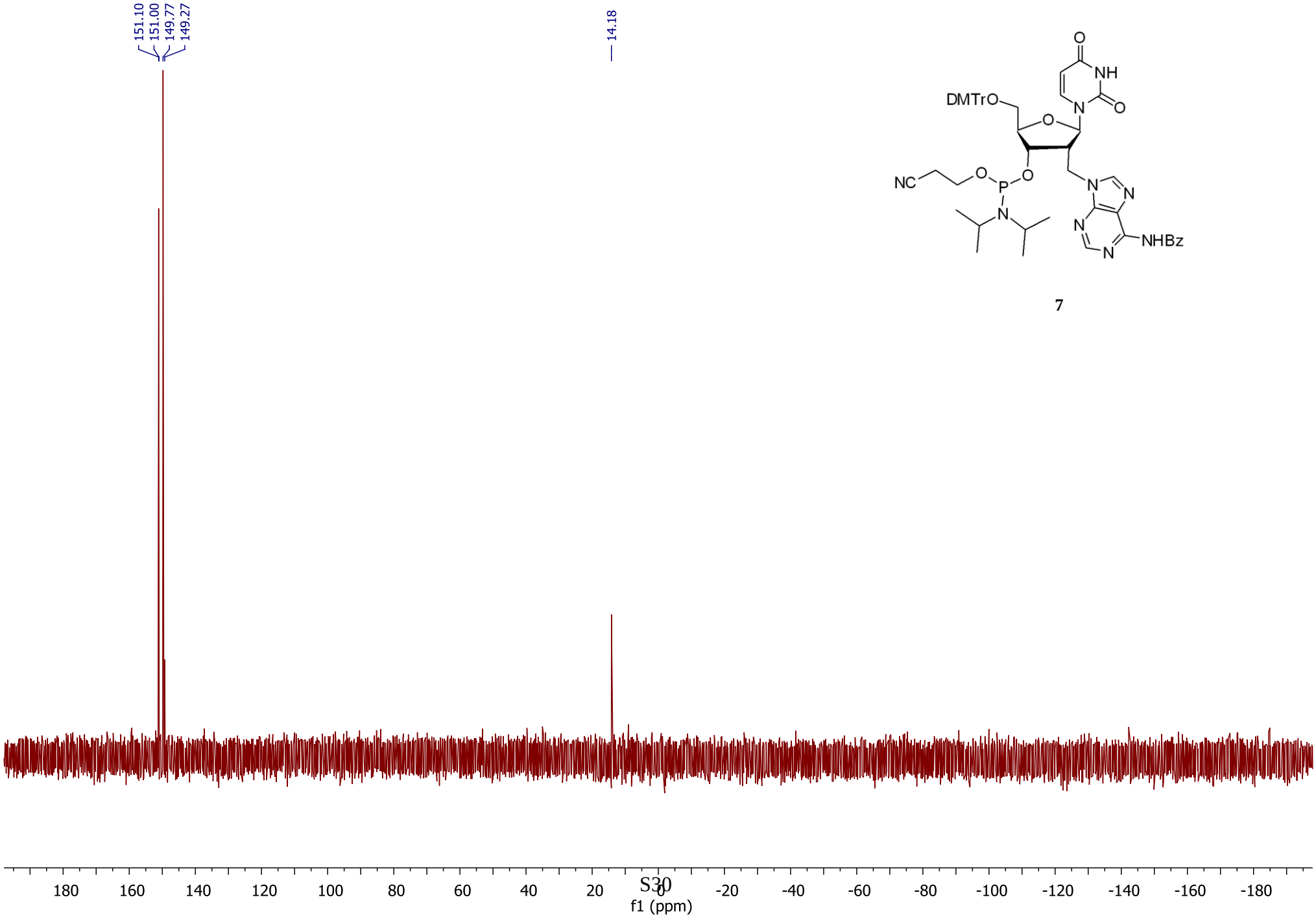


5


10.24
9.46
8.78
8.27
8.02
8.00
7.23
7.15
7.14
7.13
7.12
6.76
6.75
6.74
6.22
6.20
5.89
5.38
5.36
4.66
4.66
4.63
4.52
4.24
4.00
3.73
3.49
3.47
3.36
3.35
3.33
2.80
2.71
DMTrO
O
NH
N
O
O
HO
N
N
N
N
NHBz
6
11.5
10.5
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2
f1 (ppm)


165.2
163.1
158.9
152.4
151.8
150.3
144.2
135.1
130.2
130.1
128.9
128.3
128.2
128.1
128.0
127.3
123.1
113.4
103.2
87.4
86.3
86.2
72.2
64.0
55.4
52.0
39.8
DMTrO
O
NH
N
O
O
HO
N
N
N
N
NHBz
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)

6


151.10
151.00
149.77
149.27
14.18
DMTrO
NC
O
NH
N
O
P
NHBz
180
160
140
120
100
80
60
40
20
0
-20
-40
-60
-80
-100
-120
-140
-160
-180
f1 (ppm)

7